ASSESSING STUDENT LEARNING ONLINE: OVERCOMING RELIABILITY ISSUES

Stephen D. Arnold
The University of Arizona South
Tucson, Arizona USA

ABSTRACT

Assessing students in online university courses poses challenges to the reliability factor of the measures being utilized. Some programs have the latitude to incorporate proctored assessments, but this is not always practical in asynchronously structured courses reaching out across a broad geographic region. This paper explores digital audio and video outputs as viable student pontification alternatives.

KEYWORDS

Digital, video, online, assessment, multimedia, reliability.

1. INTRODUCTION

Assessment is the process of collecting, synthesizing, and interpreting information in order to make a decision (Airasian & Russell, 2007). Value in any instructional system comes from assessment: what is assessed in a course or a program is generally associated with value; what is valued becomes the focus of activity (Swan et al., 2007). Effective assessment typically includes ongoing "formative assessment" checkpoints, and end of term "summative assessment." Instructors signal what knowledge skills and behaviors they believe are most important by assessing them, while students quickly respond by focusing their learning accordingly (Swan et al., 2007). The end-of-course assessment method, and more specifically the requirements that underlie this assessment mode, make a difference (Struyven et al., 2006). In an online, asynchronous course, whereby the students and instructor do not meet, obtaining reliable assessment measures becomes more difficult than in a traditional face-to-face class. If exam proctoring is not an option other innovative means of measuring student learning outcome must be examined. This paper explores the creation of digital audio and video as a means for students to demonstrate learning outcomes in online university courses when exam proctoring is not possible. Although it is possible as an instructor to elicit online quizzes, papers, and projects from students it is important to collect several pieces of information about the performance being assessed to increase reliability (Airasian & Russell, 2007). Without the exam proctor possibility, determining who is (and how many are) involved in the submission of the common assessment items is problematic. As a means to strengthen assessment reliability, and foster students' creative engagement, the use of digital audio and video pontifications in online courses at a large USA university will be examined and discussed as a viable means to foster more reliable assessment outcomes for students and instructors.

2. ASSESSMENT

In any course measuring the growth students have made towards the objectives is critical to determine the effectiveness of instruction. As instructors we have to be certain that our efforts are resulting in optimum outcomes for students. In higher education written exams are often the assessment means of choice due to large numbers of students and limited instructor time. Final course grades and exams are the most common

measures of learning outcomes for seniors across Majors (NSSE, 2010). Scheduled test events tend to increase students' study time efficiency (McKenzie, 1979).

2.1 Types

There are many categorizations of assessment. First it is possible to make a distinction between assessment and test. The first being the process of determining the learning gains, and the second being the instrument or measurement tool for gathering the data. Some use the terms evaluation and assessment in a similar manner. It is important to consider the time-based snapshots of learning which are addressed by on-going incremental measurements (formative), and culminating end of term measurement of the entire term (summative). When looking at the instrument for gathering measures of learning itself there are two achievement test output categories to consider: written (conceptual, selected response), and performance (applied, task-oriented). Constructed response tests (short answer, essay), on the other hand, blur the boundaries between written and performance. They are written, and sometimes performance. "Like stories, reports, or show-your-work problems, essays and extended-response test items are important forms of performance assessments (Airasian & Russell, 2007). Educators' perspective on what constitutes performance varies, but the existence of three common characteristics prevails in identifying assessment as performance: (a) multiple evaluative criteria; (b) pre-specified quality standards; and (c) judgmental appraisal (Popham, 2011). Bloom's taxonomy of instructional objectives correlates closely to performance when students are expected to recall and apply information in an actual task. In some cases schools are requiring student exhibitions, culminating projects, experiments, solving of realistic math problems, and various other demonstrations of competence (Slavin, 2012).

2.2 Online Options

In online courses the assessment options are impacted by the student population demographics. Although students take online courses frequently at the campus of which they are a resident due to schedule restraints, remote students pose the greatest assessment challenge. Meeting with students online in videoconferencing is one way instructors assess students informally through interactive dialogue. Oftentimes students indicate dissatisfaction when instructors of online courses offer them synchronously at scheduled times due to their time/place-bound circumstances. If a student is beyond a reasonable commuting distance or has set hours of employment it is difficult to attend any scheduled class whether face-to-face or online. In the asynchronous delivery scenario the time can be more forgiving, but the options for assessment are more limited. It is possible to develop selected response (multiple-choice, true/false) and constructed response tests in course management systems, although reliability is an issue.

2.3 Reliability

When circumstances allow having a test proctored for online students will increase its reliability. Reliability in this instance refers to consistency over time. If an instructor were able to administer a test to the same student at different times ideally the results would be the same. When a test is administered, aspects related to the test construction itself, the student, graders, and various circumstances surrounding its administration could cause the results to be inconsistent (Slavin, 2012). One major factor that can affect the reliability of a non-proctored online exam is its equivalence to a take-home or open book test. In a face to face or proctored scenario the test-taker is being monitored. With the take-home scenario used in standard face-to-face classes it is possible to have supplemental in-person exams for context. Furthermore, the instructor has constant in-person engagement with the students which can provide an opportunity for oral engagement on the subject matter. In the fully online situation, however, it is difficult to gauge who or how many are working on the same exam. Given such test reliability is compromised. In one study (Watson & Sottile, 2010) students reported equitable rates of cheating in face-to-face classes as compared to online classes, but 5.2% more had someone else give them answers during an online class quiz or test. It is possible students perceive this as acceptable collaboration, not cheating.

3. THE COURSE

The events of this study transpired in a semester-based Technology for Educators course. It is organized in accordance with the Quality Matters guidelines for online instruction, and by particular themes: (a) Getting started; (b) The Internet; (c) Computer-based software; (d) Audio; (e) Interactive hypermedia; and (f) Multimedia.

3.1 Participants

In the spirit of action research "to address an actual problem in an educational setting….practical issues that will have immediate benefits for education" (Creswell, 2008) an online Technology for Educators class was analyzed in progress to understand assessment options in the online class scenario better. The course is required for all pre-service teachers in the Elementary Education program of a large southern USA university. One online section (N=21) was analyzed in-depth by the researcher within the usual scope of its delivery. The students are comprised of non-traditional older students, primarily female, and 24% Hispanic. The Elementary Education Program is a transfer program; therefore students join the program in their third year of a bachelor's degree, and with a foundational development in elementary education.

3.2 Procedures

As the basis of the content in this paper, the researcher analyzed methods for students to demonstrate competency of learning objective while increasing the reliability of assessment. The analysis is based upon the formative and summative project outcomes of students in the undergraduate Technology for Educators courses in a teacher education program. For the final exam students are given a choice of either writing a paper, or developing a comprehensive video. If they choose the video they may create it as either an individual or group project (self-selected groups). There are also other key assignments throughout the semester that have students expound upon their growth in the course content through various other digital outputs that incorporate text, static images, audio, video, or a combination. Group projects have the means to provide increased student understanding of content and instructor related advantages including multiple perspectives and pooled efforts (Young & Henquinet, 2000). From an instructor's standpoint, presentations provide an alternative means for students to demonstrate their competency vested in a culminating course project (Arnold, 2010). As a means to capture the "presentation" component of a face-to-face class in the online course medium the final project had the presentation element at its core. The process of presenting acts as reinforcement for learning that will oftentimes motivate presenters towards adequate preparation and information grounding (Arnold, 2010). Students are able to demonstrate meaningful, multidimensional tasks via this authentic assessment (Montgomery, 2002). The course at the heart of this study, Technology for Educators, was broken down into five modules, each with three weeks devoted to a certain theme. Given that the course is primarily for pre-service teachers the focus was on pedagogy and using technology to support the standards-based subjects in the classroom. Theoretically and practically, teaching requires substantive merging of content, pedagogy, and technology knowledge (Roblyer & Doering, 2010). Each of the modules had an overlying technology-based theme with multiple technologies addressed, substantial readings, academic content standards tie-in, pedagogical foundation, and emphasis on integration.

3.3 Technology Integration

Over time computer programs and online applications wax and wane. Many withstand the test of time, but undergo significant repackaging. Given the continual transitioning of computer technology the focus of the Technology for Educator courses discussed in this paper has been on helping students become more confident in exploring technology while focusing on "why" we need and choose technologies over "what" and "how." For instance the first module's theme is "online." In this theme students explore the topics digital information integrity, information literacy and multiple issues pertaining to online information exchange: legalities, safety, equity, societal, and reliability. During each module students use and create comprehensive projects with multiple cloud and computer based technologies while exploring an instructional

delivery/e-learning concept such as podcasting. Their culminating project for this module, as an example, would result in a precise public service announcement embedded in a Wiki page and uploaded to an RSS feed site (iTunes U), which demonstrates students' growth (or lack of) from the multiple readings (textbook & journal articles), instructor created multimedia presentations, and digital technology explorations. At each step of the way students are reminded to critically analyze through a K-12 teacher's lens.

3.4 Input and Output

It is possible to look at the process of developing one's content knowledge as "Input" and demonstrating what one knows as "output." More commonly this is referred to as learning and assessment. Hunter (2004) equated the terms with input of information into the students' cognitive learning processes, and output of information in a mastery of the learning objective sense so that proper assessment may occur. Output is also associated with the active process of learning whereby the process of output draws heavily upon the content knowledge students experienced through the input process reinforcing the learning (Arnold & Moshchenko, 2009). When students are given the opportunity to produce a tangible product or demonstrate something to an audience, their willingness to put forth quality increases (McTighe, 1996). Given the complexities of reliable assessment in the online environment, and that the students of the Technology for Educators course were taught the skills to create podcasts and video-based digital stories alternate assessments termed *Technology Integration Pontifications* were developed to analyze student growth in the course subject-matter.

3.5 Performance Pontification

Digital video editing is well suited for providing authentic, meaningful reflective experiences for teachers (Calandra et al., 2009). If it is more pointed in its output with specific criteria then it becomes a viable assessment tool. When constructed by the students who are being assessed, and as participants in the video, the instructor will be able to analyze the video for key levels of pontification pertaining to the course and assessment objectives as the following pontification assignment summary illustrates.

> In this final project you will create a fun digital story on a subject of interest that would be covered in an elementary classroom and that supports a content area. It may be anything from science (rocket propulsion for instance) to a social message (wash your hands frequently to reduce germs), or any other subject you would expect students of your favorite grade-level to learn (look in the content standards for a grade-level and subject of choice to identify a specific Performance Objective). In the spirit of the "reality" TV mash-ups (i.e., "Survivor" for instance where the program shows the tribes in action then cuts away to an individual sharing his/her perspective on that action in an interview scenario) I would like you to intersperse yourself into the video as the teacher giving your perspective regarding the use of technology in the learning/teaching process & with your chosen subject, while teaching the viewer about a chosen topic (i.e., the earth's rotation/tilt & seasons). Have fun with this and get kids involved if possible! You may do this as a group project with other students. (S. Arnold, personal communication, April 23, 2012)

In order to discern specific concepts critical to the outcomes of the course further detailed criteria were included. Some pertained to the technology skills whereas others were targeting educational technology and integration with the elementary school subject-matter concepts. The following are abbreviated samples of measurable objectives included in the project:

- Refer to and include specific Educational Technology supportive content from at least 6 journal article or textbook sources that were assigned during the term
- Devote about 1/3 of your video to talking about integrating technology into the classroom, and the remainder to teaching about a specific topic in a grade-level and subject of choice. Be sure to combine them so it does not appear like two separate videos.
- Include at least 3 separate motion video clips of yourself talking about integrating technology into teaching & learning.
- Include at least 2 separate audio clips of yourself talking about images, 3rd party motion video clips, technology integration, or explaining visual examples of the subject matter.
- Connect with and identify multiple standards: Information literacy, NETS*T, state academic content standards, and state educational technology standards for students.

- Make the presence of each group member equitable and evident throughout the video
- Demonstrates competency with the following technologies/technology processes: Movie Maker, Audacity, online file conversion, iTunes, YouTube, ID Tag Editor, synchronized and overlapping soundtrack and narrations.
- Effective integration of still images, motion video, text slides, overlays, soundtracks, and narrations.
- Important elements of a presentation: Introduction, body, and conclusion

On a smaller scale, and in a similar manner students were given a reading response assignment whereby they had to create an audio narrated hypermedia presentation (PowerPoint) in which they identified key points made in the readings as text on the slides and discussed them in audio format. Having met with each student individually at the beginning of the semester in a videoconferencing site, and requiring students to post audio introductions in their e-portfolios the instructor was familiar with students' voices. Given such it resulted in more personable assessment than written papers. A couple of drawbacks, especially pertaining to use in the online environment, include devising a systematically reliable video evaluation means, and the technological requirements for developing a video that represents one's content development. Validity is also important to this type of assessment; does the video allow an instructor to measure the conceptual knowledge that needs to be measured? Identifying the expected outcomes wasn't really a problem, but some students opted to read from scripts which can leave the evaluator wondering if the presenter is engaged or simply reading information that is not internalized.

3.6 Evaluating Audio & Video

From the grading perspective there is still a disparity in reliability from one instructor to another. Multiple teachers grading the same essay paper will assign grades ranging from *A* to *F* with some teachers making few to no comments or marks on the papers, but instead just producing a grade (Brimi, 2011). As noted above there is a convergence of technology use skills, technology integration with subject-matter propensity, and any given number of subtopics pertaining to educational technology covered in the course that must be weighed when evaluating a video produced by students of a Technology for Educators course. Time is a critical element in the analysis, but quality is the most decisive in determining if students are pontificating about the concepts covered in the course. In the scope of this analysis most students were able to expound upon their chosen topic (a science lesson on volcanoes for instance), but most commonly underdeveloped their connection of the lesson topic to their use of technology to demonstrate it, or other examples of technology that would further support the teaching of the lesson. The next confounding factor that tended to affect students results, whether audio or multimedia, was the technology medium being used for the output.

3.7 Technological Factors

Early in the course the technology skillsets were more limiting to the quality of course concept-infused outputs than later in the term. Given such the course was structured with less complex technological components in the beginning; create a Wiki site via a Web content management system for instance. Week-by-week new technologies are introduced. During the first three-week module students are asked to record and embed a simple wave file into the Wiki, convert a Wav to Mp3 via a "free" online media converter, and create a simple Google presentation to embed. During subsequent modules as students' efficacy climbs they are directed toward more complex technological developments such as multi-tracked audio and video outputs using programs that balance user friendliness, effectiveness, and relatively free availability. These are characteristics that are likely to encourage pre-service teachers to continue using technologies adequately when they transition to in-service status; a time that has many reeling from the steep learning and time commitment curve common during the first two years. Toward the end of the term students in the Technology for Educators course are pushing some of the low-end technologies to their limit, which inevitably impacts their output and perception of technology.

3.8 Student Media Preference

When students were asked if they preferred demonstrating content they've learned in the course through audio or video output (reading responses, interactive hypermedia, videos, etc.) over writing a paper on the same 81% strongly agreed, and 19% agreed. Given the number of technical glitches that were communicated during the term it is curious that no students indicated preference for writing a pontification paper over creating the video. Perhaps as Roblyer and Doering (2010) point out technology can improve student motivation, attitude, and interest in learning. In an end of course improvement evaluation the instructor administers to students a rank-order question indicated that students prefer audio and video enriched technologies (see Fig. 1). The question only analyzed the larger project outputs without sub-analysis of the smaller technologies that most often fed into the larger projects. In the ranks for each project/media type students identified Podcasting/Audacity (88) as their optimum output medium with E-Learning/PPT (81) and Video Pontification/Movie Maker (77) close behind in that order. In discussions between the instructor and students most seemed more enthusiastic about the outcome of their E-Learning and video activities, but the researcher believes the higher rate of technology glitches and higher complexity level of the assignments associated with their development led to lower ranking than the Podcast. Large file sizes, program freeze-ups, conversion to an iTunesU compatible mp4 file format, and student self-consciousness about presenting in the video were common concerns voiced by students during the latter part of the term devoted to the multimedia projects.

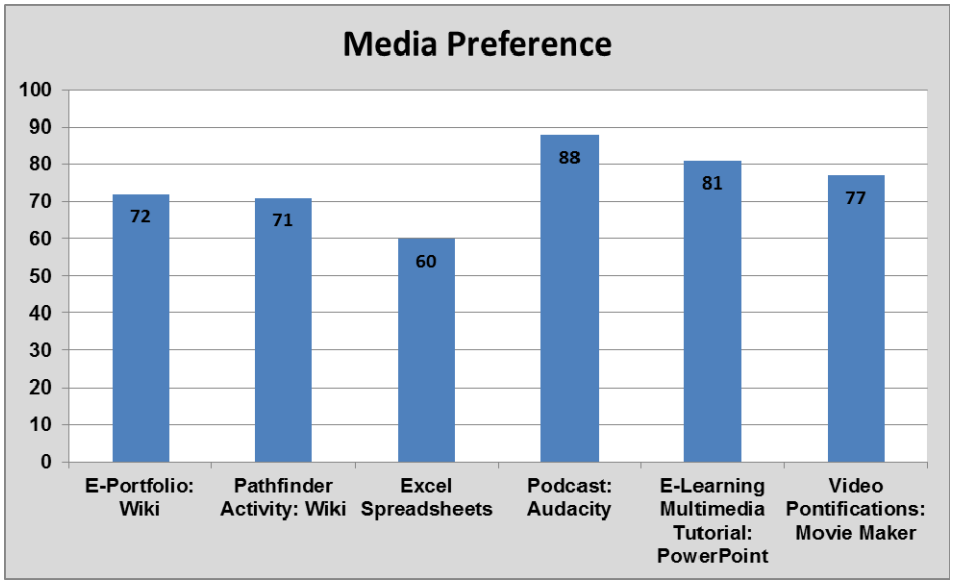

Figure 1. Class rank-order preference of major projects

3.9 Student TPACK Perception

Students (N=21) were given an additional questionnaire, the Technological Pedagogical Content Knowledge Alignment Perception Scale (TPACKAPS) at the beginning of the semester, and again at the end. Each student was asked to rate various components of the course (readings, discussions, papers, and media) on a scale from 1 to 10 for technology, pedagogy, and content knowledge emphasis (1 = none; 10 = primarily). Paired-samples t tests were conducted to determine whether students' perception of the course components varied upon completion of the course. The results indicated that the pre and post means (Table 1) varied significantly at the $p < .01$ level in students' perceptions of videos and podcasts for technology, pedagogy, and content knowledge. The positive correlation indicates that students perceive more technology focus in the beginning, but they perceive more pedagogy and content knowledge in the media projects. When asked if they felt that media output represented and demonstrated their level of learning in the course with regards to

technology, pedagogy, and content knowledge 76% strongly agreed, 14% agreed, and 10% neither agreed nor disagreed.

Table 1. Student TPACK perception means

	Video		Podcast	
	Pre	Post	Pre	Post
Technology	8.67	7.81	8.86	7.62
Pedagogy	3.95	7.91	3.95	8.24
Content Knowledge	4.86	8.48	5.14	8.10

4. EDUCATIONAL SIGNIFICANCE

The results of this study indicate that students prefer multimedia over other types of technology, view digital video and audio as TPACK rich media capable of demonstrating their competencies. Given that the course utilized in this study is heavily infused with large doses of pedagogy and content knowledge instruction in addition to the technology literacy skill development a balanced TPACK approach is modeled for the students. Furthermore, students are challenged to create outputs that equitably merge each TPACK component.

Aside from an instructional and learning tool video has been around for many years as a formative assessment, feedback, and planning tool. Common uses in this realm have included recording oneself giving a speech or presenting a student teaching lesson, real-time and post-game sports analysis, diagnosis of medical conditions or behaviors, pre and posttest analysis of research subjects, and anything that requires a comparative stop-action, archival capability. From an instructor's point of view the digital audio and video output can offer a creative and visual dimension not offered in print. From the student perspective whereby they are interjecting an audible or visual presence in the media it is typically more customary as a self-assessment tool. Video tools are not uncommon as a means to teach content to others, even with the self in the visual mix. Through the use of audio and video students are able to solidify their learning due to the increased cognitive processing needed to develop quality output. In addition students will encounter added motivation due to prospect of having a novel means to demonstrate their competency. As a demonstration of what one has learned based upon engagement in a course, however, digital audio and video in the online class environment is an under-utilized and viable output. It warrants further analysis as an output tool not only in Educational Technology, but in less technology-focused disciplines as well.

REFERENCES

Airasian, P.W., & Russell, M.K. (2007). *Classroom assessment: Concepts and application* (6th ed.). New York, NY: McGraw-Hill.

Arnold, S.D. (2010). Student e-cooperative: Computer mediated group projects. In the *International Technology, Education, and Development Conference Proceedings* CD (ISBN: 978-84-613-5538-9). Valencia, Spain: INTED.

Arnold, S.D., & Moshchenko, M. (2009). Technology input versus input and output: Does it result in learning differences among elementary school students? In C.D. Maddux (Ed.), *Research Highlights in Technology and Teacher Education 2009* (pp. 1-9). Chesapeake, VA: Society for Information Technology & Teacher Education.

Brimi, H.M. (2011). Reliability of grading high school work in English. *Practical Assessment, Research & Evaluation*, 16(17), pp. 1-12.

Calandra, B., Brantley-Dias, L., Lee, J.K., & Fox, D.L. (2009). Using video editing to cultivate novice teachers' practice. *Journal of Research on Technology in Education*, 42 (1), pp. 73-94.

Creswell, J.W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Hunter, R. (2004). *Madeline Hunter's mastery teaching: Increasing instructional effectiveness in elementary and secondary schools.* Thousand Oaks, CA: Corwin Press.

McKenzie, G.R. (1979). Effects of questions and test-like events on achievement and on-task behavior in a classroom concept learning presentation. *Journal of Educational Research*, 72(6), pp. 348-351.

McTighe, J. (1996). What happens between assessments? *Educational Leadership*, 54 (4), pp. 6-12.

Montgomery, K. (2002). Authentic tasks and rubrics: Going beyond traditional assessments in college teaching. *College Teaching*, 50 (1), pp. 34-39.

National Survey of Student Engagement [NSSE]. (2010). *Major differences: Examining student engagement by field of study—annual results 2010*. Bloomington, IN: Indiana University Center for Postsecondary Research.

Popham, J.W. (2011). *Classroom assessment* (6th ed.). Boston, MA: Pearson Education, Inc.

Roblyer, M.D., & Doering, A.H. (2010). *Integrating educational technology into teaching* (5th ed.). Boston, MA: Pearson Education, Inc.

Slavin, R.E. (2012). *Educational psychology: Theory and practice* (10th ed.). Boston, MA: Pearson Education, Inc.

Struyven, K., Dochy, F., Janssens, S., Schelfhout, W., & Gielen, S. (2006). The Overall Effects of End-of-Course Assessment on Student Performance: A Comparison Between Multiple Choice Testing, Peer Assessment, Case-Based Assessment and Portfolio Assessment. *Studies in Educational Evaluation*, 32, pp. 202-222.

Swan, K., Schenker, J., Arnold, S. & Kuo, C-L. (2007). Shaping online discussion: Assessment matters. *E-mentor*, 1 (18), at: http://www.e-mentor.edu.pl/_xml/wydania/18/390.pdf

Watson, G., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses? *Online Journal of Distance Administration*, 13 (1). Retrieved from http://www.westga.edu/~distance/ojdla/spring131/watson131.html

Young, C.B., & Henquinet, J.A. (2000). A conceptual framework for designing group projects. *Journal of Education for Business*, 76 (1), pp. 56-60.